Comparing MLR and ANN models for school building electrical energy prediction in Osijek-Baranja County in Croatia

Begić Juričić, Hana; Krstić, Hrvoje

Source / Izvornik: **Energy reports, 2024, 12, 3595 - 3606**

Journal article, Published version
Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

https://doi.org/10.1016/j.egyr.2024.09.039

Energy Reports 12 (2024) 3595–3606

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research Paper

Comparing MLR and ANN models for school building electrical energy prediction in Osijek-Baranja County in Croatia

Hana Begić Juričić *, Hrvoje Krstić

Faculty of Civil Engineering and Architecture Osijek, Josip Juraj Strossmayer University of Osijek, Croatia

ARTICLE INFO

Keywords:
School buildings
Energy consumption
Electrical energy
Energy prediction

ABSTRACT

This paper presents a study conducted in Osijek-Baranja County, Croatia, to predict electrical energy consumption in school buildings. Data from the Energy Management Information System (EMIS) database for primary and secondary schools were analyzed using Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN). The ANN model achieved a high R^2 of 0957 in the training set, with lower Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) than the MLR model, which had an R^2 of 0950. On the validation set, the ANN model maintained an R^2 of 0954 and showed slightly better performance with a lower Coefficient of Variation of RMSE (CVRMSE) of 19,79 %, compared to the MLR model's CVRMSE of 20,50 %. These results indicate that the ANN model generally provides more accurate and reliable predictions for energy consumption in school buildings. However, both models provided a robust positive correlation between the predicted and actual values.

1. Introduction

Buildings have been found to account for more than 40 % of energy consumption in many countries worldwide (Alghoul et al., 2016; Cao et al., 2016; D'Agostino et al., 2017). This is also true for Croatia, where, according to the Energy in Croatia 2022 report, the total energy consumption in buildings accounts for over 47 % of the final energy consumption (Fig. 1) (Republic of Croatia Ministry of Economy and Sustainable Development, 2022).

Public buildings, especially schools, offer a great opportunity for establishing energy-efficient measures. This is because they have a large proportion of the total number of buildings, which leads to a substantial impact on both the amount of energy used and the financial resources of the country (Dimoudi and Kostarela, 2009). These facilities play a large role in water and energy consumption. However, despite the potential for huge savings through the use of energy and water-efficient technology, there are several challenges to their adoption that are more severe than those in other sectors (Papadakis and Katsaprakakis, 2023; Bertone et al., 2018). Regarding Croatia's building stock composition, the Croatian Long-Term Strategy for National Building Stock Renovation by 2050 offers valuable insights. The data reveals that in 2018, non-residential buildings constituted a significant share, encompassing 28 % of the total building area. Notably, educational buildings within this category hold an exceptionally high proportion, accounting for 10 % of the entire non-residential building area. This dominance of educational buildings is expected to persist, with projections indicating a consistent 10 % share in 2030, 2040, and even 2050 (Table 1) (Ministarstvo prostornog uređenja et al., 2020).

Moreover, the breakdown of building stock composition, particularly the prominence of non-residential buildings, sheds light on specific areas that require attention for energy efficiency improvements. The projection that educational buildings will maintain their significant share in the future highlights the long-term implications of energy consumption patterns. It suggests that implementing sustainable solutions will have a lasting impact on reducing energy consumption and mitigating environmental impacts in Croatia. Energy-saving retrofits for educational buildings have become increasingly important in recent years (Papadakis and Katsaprakakis, 2023; te Kulve et al., 2022; Araújo et al., 2023; Teli et al., 2017; Zapata-Lancaster et al., 2023). Schools have a significant societal obligation in the construction industry because of their role in education (Lizana et al., 2018). According to the literature, the school's energy expenditures are the second highest expense for the school, after the salaries of teachers and personnel (Pereira et al., 2014). Additionally, the Energy in Croatia report from 2022 highlights that in 2022, the share of electrical energy consumption in the final energy consumption by fuel equals a significant 20,8 %, right behind petroleum products (Fig. 2).

* Correspondence to: Vladimir Prelog Street 3, Osijek 31000, Croatia.
E-mail addresses: hbegic@gfos.hr (H. Begić Juričić), hrvoje.krstic@gfos.hr (H. Krstić).

https://doi.org/10.1016/j.egyr.2024.09.039
Received 21 March 2024; Received in revised form 20 August 2024; Accepted 16 September 2024
Available online 27 September 2024

Nomenclature

AEC	Average annual electrical energy consumption, kWh/year
A_k	Total useful surface area, m^2
R^2	Coefficient of determination
TNU	Total number of users (employees and pupils), number of users
V_e	Heated volume of the building, m^3

List of abbreviations

ANN	Artificial neural network
AE	Autoencoders
BIM	Building information modeling
CART	Classification and regression tree
CNN	Convolutional neural network
CVRMSE	Coefficient of variation of the root mean square error
EMIS	Energy Management Information System
GBR	Gradient boosting regressor
HVAC	Heating, ventilation, and air conditioning
kWh	Kilowatt-hour
LSTM	Long short-term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MLP	Multi-layer perceptron
MLR	Multiple linear regression
MSE	Mean square error
nRMSE	Normalized root mean square error
QC	Quantum computing
RF	Random forest
RMSE	Root- mean-square error
RNN	Recurrent neural network
SARIMA	Seasonal Autoregressive Integrated Moving Average

Many different elements influence the energy consumption of a school building (Pereira et al., 2014; Katafygiotou and Serghides, 2014; Perez and Capeluto, 2009). Hence, it is crucial to identify these elements while constructing a prediction model for energy consumption. Anticipating energy consumption in constructed facilities is crucial for grid management to conserve power, ensure optimal usage, and prevent wastage (Bhowmik et al., 2017; Chen et al., 2023). Nevertheless, making precise predictions is challenging because of unforeseeable circumstances and the overall disorderliness of data, and the techniques employed frequently produce inaccurate forecasts (Mohammed et al., 2021). In this perspective, the Ordinance on Systematic Energy Management in the Public Sector (Official Gazette 18/2015) was adopted in 2015, which prescribes the obligation to manage energy and water consumption, analysis of consumption, the method of reporting on energy and water consumption, and the methodology of systematic energy management in the public sector (Ministarstvo graditeljstva i prostornoga uređenja, 2015). Also, the Energy Management Information System (EMIS) was developed to provide support for strategic planning of energy and sustainable management of energy resources in buildings that are owned or used by cities, counties, the Government of the Republic of Croatia, as well as in buildings of other government budgetary and extra-budgetary users, and public authorities (Agencija za pravni promet i posredovanje nekretninama, 2015). In this context, only one research has been conducted in Croatia regarding analyzing the performance and buildings characteristics obtained from EMIS (Krstić and Teni, 2018).

Table 1
Projection of the total area of non-residential buildings in Croatia in 2030, 2040 and 2050 (Ministarstvo prostornog uređenja et al., 2020).

Type of non-residential building	2030 [m^2]	2040 [m^2]	2050 [m^2]
Office	10 309 712	10 831 614	11 082 926
Educational	6 236 465	6 552 169	6 704 190
Hotels and restaurants	4 650 511	4 885 930	4 999 292
Hospitals	3 280 271	3 446 326	3 526 286
Sport	462 823	486 252	497 534
Stores	12 833 465	13 483 125	13 795 956
Other	24 303 780	25 534 093	26 126 528
Total area:	**62 077 026**	**65 219 509**	**66 732 712**

A review of the literature indicates substantial global research on forecasting and analyzing school energy use, with several studies conducted in neighboring countries like Serbia Herzegovina (Bećirović and Vasić, 2013; Stanković et al., 2009; Đukanović et al., 2022; Jurišević et al., 2018; Jovanović et al., 2018) and Bosnia and Herzegovina (Katicccć et al., 2020, 2021; Katicccc et al., 2021; Katicccć and Krstić, 2022). However, there is a notable research gap in Croatia, particularly concerning the prediction of electrical energy consumption in schools. This gap highlights the need for a focused study that identifies the key variables influencing energy use in Croatian schools and develops accurate prediction models that can inform energy management and policy decisions.

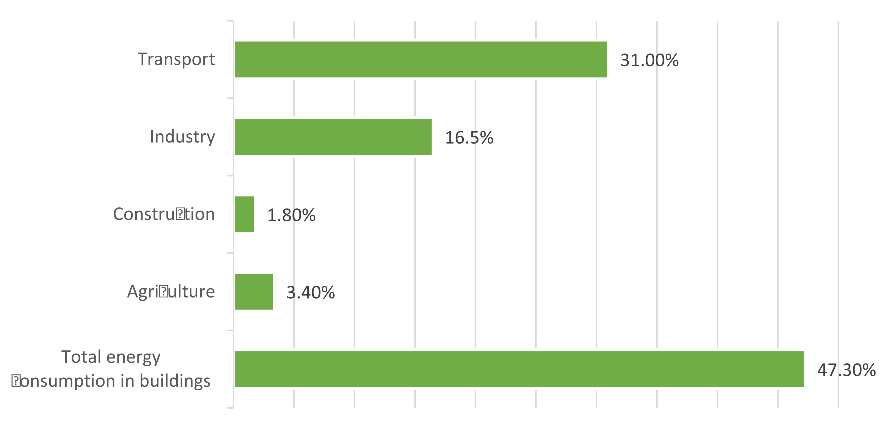

Fig. 1. Share of total consumption in buildings in 2022 in final energy consumption in Croatia (Republic of Croatia Ministry of Economy and Sustainable Development, 2022).

To address these gaps, this study aims to:

- Identify the factors influencing electrical energy consumption in school buildings in Osijek-baranja county in Croatia.
- Develop prediction models using Artificial Neural Networks (ANNs) and Multiple Linear Regression (MLR).
- Evaluate the accuracy of these models to ensure they meet the practical needs of building maintenance managers and school principals.

In light of these observations, this study aims to bridge this gap by focusing on two primary research questions:

RQ1. What are the key factors influencing electrical energy consumption in schools of Osijek-baranja county?

RQ2. How do multiple linear regression (MLR) and artificial neural networks (ANNs) perform in predicting electrical energy consumption in these schools?

The models were developed using actual school electrical energy consumption data acquired from EMIS to ensure accurate outcomes. Maintenance managers responsible for buildings and school principals could utilize the newly developed predictive model for budget planning. This study offers a straightforward, precise, and enhanced tool for advancing future electrical energy consumption predictions, facilitating more precise budget allocations. Moreover, it enhances energy management by furnishing a more precise predictive model for future electrical energy consumption.

The rest of the paper is structured as follows: in Section 2 there is a literature review, in Section 3 the applied methodology is described, Section 4 presents the developed model using MLR, and Section 5 presents the developed model using ANN. In Section 6, a comparison is performed among the developed models, and in Section 7, there is a discussion, and conclusions are drawn.

2. Literature review

The emphasis on educational buildings as a significant portion of non-residential buildings underscores the importance of targeting these structures for energy-saving initiatives (Geraldi and Ghisi, 2020; Vassallo, 2020; Børke, 2006). Some research even states that school buildings represent more than 95 % of global energy consumption (Corgnati et al., 2008). In addition, the school's energy consumption is a contributing factor to the overall operating expenses of the institution. Following the salaries of teachers and staff, energy expenditures represent the second most substantial expense (Pereira et al., 2014). Also, forecasting and predicting building energy usage is the primary objective for building energy management and facility managers (Ahmad et al., 2017). Over the past few decades, many methods have been suggested for predicting energy use in building construction. Most case studies utilize past energy consumption data to build the prediction models. The methodologies developed for predicting building energy use can be classified into two categories: statistical methods and artificial intelligence (Li et al., 2017). An extensive review of data-driven tools for building energy consumption prediction by Olu-Ajayi et al. highlights that ANN produced better performances in more studies than statistical tools such as MLR. Nevertheless, MLR exhibited optimal outcomes in particular scenarios, such as predicting annual energy use (Olu-Ajayi et al., 2023).

2.1. *Regression analysis and artificial neural networks (ANNs) in predicting energy consumption*

González and Zamarreño introduced a novel method for accurately predicting electric loads using a feedback ANN trained using a hybrid algorithm. The model's implementation provided notable results for electric load forecasting in buildings. It was stated that the new energy predictor demonstrated an accuracy similar to the best findings documented in the literature. The authors emphasize that the primary advantage of this system is its simplicity, which is derived from the straightforward nature of the proposed tool and the small and easily accessible resources required for its implementation in modern automation systems (Gonzalez and Zamarreno, 2005). Tso and Yau introduced three modeling methods, regression analysis, decision tree, and ANNs, to forecast electricity consumption. The dependent variable in this study was the aggregate weekly electrical energy consumption (kWh). They conducted a two-phase survey in the summer and winter of 1999–2000. The included independent variables were housing type, household characteristics, and appliance ownership, which were hypothesized to impact electricity energy usage. All three proposed models were determined to be comparable. During the summer phase, the decision tree model had a slight advantage over the other two techniques. On the other hand, ANN had a slightly superior performance throughout the winter phase compared to the other two alternatives (Tso and Yau,

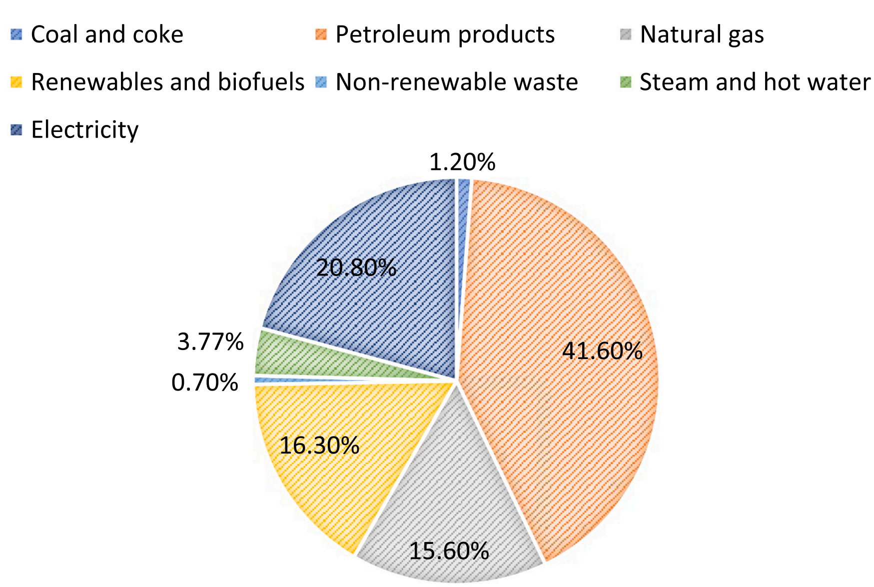

Fig. 2. Final energy consumption by energy form (Republic of Croatia Ministry of Economy and Sustainable Development, 2022).

2007). Biswas et al. demonstrated the utilization of ANNs to estimate and predict energy usage in residential structures. The input variables from the house data comprised the number of days, outdoor temperature, and solar radiation. On the other hand, the output variables were the energy consumption of the house and the heat pump. The authors affirmed that the outcomes were deemed satisfactory for the provided data set and exhibited a level of statistical analysis that was in line with previous publications (Biswas et al., 2016). Ahmad et al. conducted a comparative analysis between a feed-forward back-propagation ANN and random forest (RF) to forecast a hotel's hourly HVAC energy consumption in Madrid, Spain. The analysis revealed that the ANN demonstrated a slight performance improvement, as indicated by a root-mean-square error (RMSE) of 4,97 compared to the RMSE of 6,10 achieved using RF. Based on the findings, it was determined that both models were viable and efficient in forecasting hourly HVAC electricity usage (Ahmad et al., 2017). El Alaoui et al. compared in theri study the effectiveness of machine learning models and Seasonal Autoregressive Integrated Moving Averag (SARIMA) models in predicting heating energy usage for an administrative building in Chefchaouen City, Morocco. The results showed that while machine learning models, including artificial neural networks, generally outperform SARIMA, SARIMA models still achieve good accuracy with limited training data. The best performance was observed with the artificial neural network (R2=0.97, nRMSE=12.60 %, MAE=0.19 kWh) (El Alaoui et al., 2023). Afzal et al. presented a study that focuses on predicting cooling and heating loads using a multilayer perceptron (MLP) neural network. To optimize the MLP model, it was combined with eight meta-heuristic algorithms through a hybrid approach. The analysis revealed that the MLP-PSOGWO model outperformed others, achieving the highest accuracy, with R^2 values of 0.966 for cooling loads and 0.998 for heating loads (Afzal et al., 2024).

2.2. Energy consumption prediction in school buildings

Corgnati et al. carried out a field survey to gather, process, and examine data on the energy usage for heating purposes in a sample of 138 buildings (117 high schools, nine office buildings, and 12 residences for school keepers) in the Provincia di Torino. The results demonstrate a satisfactory correlation between the yearly measured and adjusted conventional heat supply. On the other hand, the monthly heat supply figures reveal a significant disparity between the actual and estimated energy consumption. The observed outcome can be attributed to the variability in monitoring intervals, wherein the meters may not be precisely read at the end of each month. According to the authors, this methodology is well-suited for evaluating extensive building stocks over a long period. It is beneficial for measuring the yearly and total cost of energy service (Corgnati et al., 2008). Capozzoli et al. analyzed the energy use for heating in 80 school buildings in northern Italy. Two estimating models, namely MLR and Classification and Regression Tree (CART), were created and evaluated to evaluate energy usage. The two models were evaluated based on statistical coefficients. The analysis concluded that the gross heated volume, heat transfer surfaces, boiler size, and thermal transmittance of windows primarily impacted the heating energy consumption of the school buildings under consideration (Capozzoli et al., 2015). Beusker et al. conducted an empirical study that analyzed the elements influencing the heating energy usage of municipal schools and sports facilities. The study was based on a random sample of 105 properties in Stuttgart. The study's primary aim was to propose a standardized and adaptable estimating model that can identify essential characteristics for effectively benchmarking, evaluating, and predicting the heating energy consumption of schools, sports facilities, and combinations of both types of usage. Various linear and non-linear regression models have been systematically created and evaluated to forecast the heating energy consumption of schools and sports facilities. A validation test was conducted to assess the forecast accuracy of the stated model, which included five properties. The authors have discovered that the introduced model demonstrates high accuracy and satisfies all necessary characteristics for producing reliable, impartial, and efficient estimations (Beusker et al., 2012). Authors Mohammed et al. proposed a multiple regression-based model for estimating the energy consumption of school facilities in Saudi Arabia. In order to enhance the precision and effectiveness of the feasibility study, accurate school energy consumption data was gathered and utilized to construct the model. With reliable benchmark data on energy use, the running costs of a school after its construction can be relatively inexpensive. The model provides a practical and inexpensive approach for government entities to utilize in consumption prediction (Mohammed et al., 2021). Faiq et al. proposed the utilization of a Long Short-Term Memory (LSTM) model to anticipate the energy consumption of an academic building. This model incorporates forecasted weather data. The predictive model was developed by examining the correlations between energy consumption and weather data. The authors emphasize the importance of having precise day-ahead weather forecasting parameters in order to achieve accurate predictions. Furthermore, LSTM necessitates the inclusion of external variables, such as environmental factors (e.g., temperature and wind speed) and schedule-related variables, to enhance the precision of the predictive model. The authors suggest that future research could enhance the model by incorporating additional characteristics, such as the number of occupants (Faiq et al., 2023). Cao et al. proposed a model for predicting energy consumption in educational facilities by including geographical variables into time series data. They also examined the impact of various aspects on the model using the cooperative game theory approach. The validity of the suggested model is confirmed through its application to an educational facility located in Xi'an, Shaanxi Province. The results indicate that the integrated energy consumption prediction model exhibits a reduction in RMSE value ranging from 13.64 % to 34.55 % compared to previous prediction models. Additionally, the MAE value is lowered by 10.25–30.54 %, demonstrating improved forecast accuracy (Cao et al., 2023). Shahid et al. developed a predictive model for estimating electricity and district heating usage over one or multiple days. They utilized advanced machine learning techniques such as Multivariate Recurrent Neural Network (RNN), LSTM, Convolutional Neural Networks (CNNs), and Autoencoders (AE). The model was trained using actual consumption data from six public schools in a Swedish municipality. The experimental results indicate that the model has attained a high level of accuracy, with RMSE and normalized root mean square error (nRMSE) values ranging from 18 % to 25 % and 5–6 % respectively for electricity, and from 20 % to 30 % RMSE and 5 % nRMSE for district heating (Shahid et al., 2023). Elhabyb et al. developed a predictive model for electricity consumption in three privately owned research university buildings based on a dataset gathered from January 2020 to January 2023. Developing the prediction model entailed data preparation, encompassing tasks such as addressing missing data and determining the significance of features. The researchers employed three machine learning techniques, namely gradient boosting regressor (GBR), LSTM, and RF, as the algorithms for the predictive model. As a future possibility for research, the authors propose employing more robust computational systems or platforms to execute the LSTM algorithm, which has the potential to enhance its overall performance (Elhabyb et al., 2024). Doiphode and Najafi proposed a multi-layer perceptron (MLP) neural network model to predict the monthly energy usage of K-12 schools in Brevard County, Florida. The network's inputs consist of the number of occupants, the number of working days per month, the building area, the average monthly external temperature, and the relative humidity, where the network's output is the monthly energy usage. The chosen network was effectively trained utilizing three years of energy usage data from 25 high schools, middle schools, and elementary schools. The findings demonstrated that the neural network model that was created can precisely predict the monthly energy usage of schools (Doiphode and Najafi, 2020). Geraldi et al. introduced a data-driven approach that utilizes Bayesian Networks to forecast energy use. The authors collected monthly energy bills over a three-year period from 90 public schools located in southern Brazil. They also obtained information about each school's floor-plan area, number of students, level of education, number of floors, and incidence of events. As possible directions for future research, the authors suggest enhancing the database by incorporating more features and expanding the existing dataset (Geraldi and Ghisi, 2020). Run et al. employed an MLR model to forecast the hourly electricity energy usage for school buildings in the southern region of France throughout the winter season. The analysis revealed that the coefficient of determination (R^2) for the training set is 74 %, while for the testing set it is 77 %. One limitation of this model, as pointed out by the authors, is that it underestimates outcomes for power consumption levels above 30 kWh/h. However, the model serves as a starting point for future study aimed at improving its prediction capacity (Run et al., 2023). Li et al. proposed a method for short-term prediction of air conditioning (AC) energy consumption, focusing on uncertain usage patterns influenced by occupant behavior. Applied to an educational building, the method uses cluster analysis to identify typical patterns and the weighted k-Nearest-Neighbors technique to forecast the AC usage rate. An RF model is then developed, evaluating each variable's significance. To improve accuracy in unpredictable conditions, a support vector machine is employed. The results show that while both models effectively predict AC energy consumption, the enhanced model offers notably greater accuracy under uncertain conditions (Li et al., 2023). Tariq et al. conducted a study that explores various artificial intelligence models such as decision trees, K-nearest neighbors, GBR, and LSTM to predict energy usage in schools, emphasizing the impact of factors like school size and AC capacity on annual consumption. The findings indicate that while Decision Trees perform well in training with low prediction errors, K-Nearest Neighbors struggle with overfitting. GBR and LSTM models excel in handling diverse data ranges. The research highlights the role of sustainable educational buildings as interactive learning environments that teach students about energy efficiency and sustainability, encouraging the use of AI tools to optimize energy consumption in these spaces (Tariq et al., 2024).

2.3. Emerging techniques in energy consumption prediction

Advancements in predictive modeling techniques have revolutionized the field of energy consumption prediction, offering novel approaches to enhance accuracy, efficiency, and applicability across diverse domains.

In this perspective, quantum computing (QC) is an emerging discipline that utilizes the principles of quantum mechanics to carry out information processing tasks that are not achievable with conventional computers. Quantum computers has problem-solving techniques that are inaccessible to classical computers, hence enabling them to utilize unique methodologies (Giani and Eldredge, 2021). Deng et al. proposed a novel optimization approach utilizing quantum annealing for model predictive control of a rooftop unit. In contrast to conventional optimization techniques, the authors achieved comparable solutions with less than 2 % discrepancies and enhanced computational efficiency, reducing the time required from hours to seconds. By incorporating day-ahead price time-of-use demand response signals, they achieved an impressive 80 % decrease in overall electricity usage and a 21 % decrease in electricity expenses (Deng et al., 2023). Kumar K et al. provided a comprehensive summary of the Home Energy Management System and the techniques used for load forecasting. They also introduced a new method that uses a Quantum Support Vector Machine to estimate periodic power usage accurately. The authors analyzed the energy consumption patterns of household appliances, sun irradiation, and the total load, and the model is emphasized for its notable versatility and performance (Nutakki et al., 2024).

In the realm of energy management, alongside advancements in quantum computing for optimizing computational efficiency and decision-making strategies, CNNs have emerged as a transformative technology. Xu et al. developed a building energy consumption optimization model based on CNN and BIM. The authors wanted to enhance the advancement of BIM technology by examining professional collaborative design, pipeline installation, and construction optimization methods based on BIM technology. Additionally, they investigated the viability of utilizing BIM energy consumption analysis technology to support energy-saving design applications during the initial phase of building scheme design. The authors offered three optimization approaches for the energy-saving design of the case building based on a thorough investigation of the internal light environment, building energy consumption, and construction cost (Xu and Liu, 2023). Wang et al. introduced a hybrid neural network prediction model that integrates an attention mechanism, Bidirectional Gate Recurrent Unit, Convolutional CNN, and the residual connection. The model employs a Bidirectional Gated Recurrent Unit to train the feature vectors extracted by a Convolutional Neural Network (CNN) in a two-way cycle. The study revealed that the constructed model demonstrates higher accuracy in predicting outcomes during working hours compared to non-working hours in the office building. Additionally, the model's prediction accuracy is greater for the same season compared to the entire year (Wang et al., 2023).

3. Methodology

Forecasting the energy usage of a school facility, particularly with precision, might pose difficulties. The challenge lies in numerous unpredictable variables, such as the characteristics of physical materials and usage circumstances (Mohammed et al., 2021). Moreover, the estimation of energy usage is influenced by the challenge of identifying suitable patterns that connect input variables and energy consumption, and the complexity of precisely identifying the correlation between input variables and output (Kavgic et al., 2010; Raza and Khosravi, 2015). State-of-the-art research has argued that MLR and ANN models can solve such problems. Accordingly, this study proposes two models for estimating the energy consumption of school buildings in Osijek-Baranja County, Croatia, one using MLR and one using ANN. The proposed workflow for the study is presented in Fig. 3.

3.1. Data collection

The data was obtained from the EMIS database for primary and secondary schools in the Osijek-Baranja County in Croatia. The obtained data comprised several parameters that affect energy usage, primarily related to building characteristics. These values were used as input data in the models. The models included one parameter, the average annual energy consumption measured in kWh / year, as the output. The data acquired encompassed the precise energy usage of 149 school buildings. To simplify the process, the data was arranged in Microsoft Excel workbooks (Microsoft, 2024), which is compatible with TIBCO Statistica® 14.1.0 (Cloud Software Group Inc, 2024), the software used to develop both models.

3.2. Variable selection

Firstly, it was necessary to determine the input (independent) variables affecting school electrical energy consumption. A review of previous research in this area was used to identify the most significant input variables, where the average annual energy consumption was the output. The identified significant input variables for predicting electrical energy consumption of school buildings in Osijek-Baranja County, Croatia, are presented in Table 2 together with the relevant research where they were also used.

Each of these variables was selected based on their proven impact on energy use in buildings. The total number of users (TNU) affects energy consumption through increased operational demands, while the total useful surface area (A_k) and heated volume (V_e) influence the energy

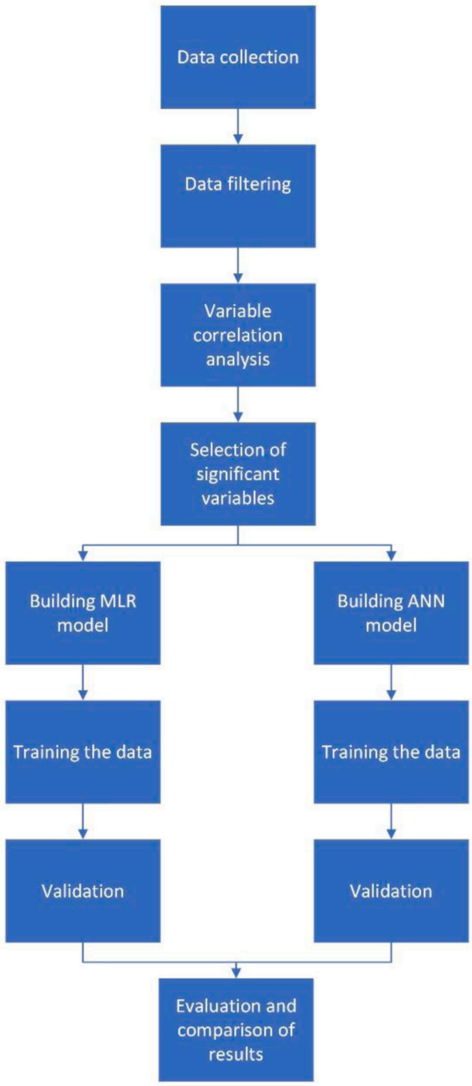

Fig. 3. Proposed workflow of the study.

required for temperature regulation. By including these variables, our models aim to provide a comprehensive understanding of the factors contributing to energy use in buildings.

Also, after determining the input variables, it was necessary to determine the output (dependent) variable, which is presented in Table 3.

Table 2
Literature review-based identified significant input variables for the development of a model for predicting the consumption of electrical energy in school buildings.

Input variable	Description	Label	Unit of measure	Also used in
Total number of users (employees and pupils)	Represents the total number of individuals using the building, affecting energy consumption.	*TNU*	number of users	(Mohammed et al., 2021; Tso and Yau, 2007; Raatikainen et al., 2016; Alshibani, 2020; Kim et al., 2019; Desideri and Proietti, 2002; Gaitani et al., 2010)
Total useful surface area	The total area within the building that is heated/cooled, influencing energy requirements.	A_k	m^2	(Mohammed et al., 2021; Tso and Yau, 2007; Raatikainen et al., 2016; Alshibani, 2020; Kim et al., 2019; Gaitani et al., 2010; Issa et al., 2010)
Heated volume of the building	The total volume of space within the building that requires heating, impacting energy use.	V_e	m^3	(Corgnati et al., 2008; Raatikainen et al., 2016; Desideri and Proietti, 2002)

Table 3
Output variable for the development of a model for predicting the consumption of electrical energy in school buildings.

Output variable	Label	Unit of measure
Average annual electrical energy consumption	*AEC*	kWh/year

3.3. Data analysis

Table 4 presents the basic statistical measures of the gathered data, revealing the data's quality and detecting any false information, if present. The table displays different school areas, volumes, and numbers of users, enabling the creation of an accurate model that encompasses a wide range of scenarios and circumstances.

The provided statistics offer a comprehensive view of the distribution of each variable:

- *TNU*: Ranging from 3 to 894, with an average of approximately 132,96 and a notable standard deviation of 176,21 indicating considerable variability in the number of users among the sample.
- A_k: With values spanning from 37,70 to 6210,26, the total area is notably higher at around 1107,71 accompanied by a substantial standard deviation of 1316,78 indicating significant variance in A_k across the dataset.
- V_e: Demonstrating a wide range of values from 113,10 to 25935,40 the mean heated volume stands at 4435,26 with a considerable standard deviation of 5357,29 indicating notable disparities in V_e among the observations.
- *AEC*: Spanning from 393,3855 to 59566,18 with an average consumption of 13624,54 and a substantial standard deviation of 15486,72 reflecting significant diversity in *AEC* across the sampled school buildings.

The presence of small district schools and larger main ones likely influences the observed variability in electrical energy consumption across the dataset. Such differences in the scale and infrastructure of educational facilities can lead to significant variations in energy usage. Small district schools may have fewer students, smaller buildings, and less complex HVAC systems than larger main schools. Consequently, their energy consumption levels are expected to be lower on average. Conversely, larger main schools have more extensive facilities, including larger classrooms and administrative areas requiring more energy. Thus, the wide range of electrical energy consumption values observed in the dataset is consistent with the differing characteristics and sizes of the schools included, highlighting the importance of considering such factors when analyzing energy usage patterns and developing predictive models.

3.4. Variable correlations

Table 5 presents a correlation matrix showing the relationships between different input variables within the studied context. Each cell in

Table 4
Statistical analysis of the data.

Type of variable	Variable	Valid N	Mean	Minimum	Maximum	St. Deviation
Input	*TNU*	149	132,96	3,0000	894,00	176,21
	A_k	149	1107,71	37,7000	6210,26	1316,78
	V_e	149	4435,26	113,1000	25935,40	5357,29
Output	*AEC*	149	13624,54	393,3855	59566,18	15486,72

Table 5
Correlation of model input variables.

Variable	*TNU*	A_k	V_e
TNU	1,000000	0,771756	0,773813
A_k	0,771756	1,000000	0,847632
V_e	0,773813	0,847632	1,000000

the table displays the correlation coefficient between two variables, indicating the strength and direction of their relationship.

The diagonal elements of the matrix (where variables are compared with themselves) are always 1, as they represent the correlation of a variable with itself, which is perfect and expected. The off-diagonal elements reveal the correlations between pairs of variables. For example:

- The correlation coefficient between *TNU* and A_k is 0,771756, indicating a strong positive relationship between the total number of users and the building area.
- Similarly, the correlation coefficient between *TNU* and V_e is 0,773813, indicating a strong positive relationship between the total number of users and the building's heated volume.
- The correlation coefficient between V_e and A_k is 0,847632, indicating an even stronger positive relationship between the building volume and area.

It is essential to observe that a positive value signifies a positive correlation, indicating that an increase in one variable results in an increase in the other (Schober et al., 2018). Conversely, a negative value signifies a negative correlation, indicating that an increase in one variable will decrease the other. However, in this case, there are no such relationships. All of the correlations are very significant, as indicated by their high values and the red markings. These noted correlations have a significance level of $p<0.05$. In addition to analyzing the correlation among the input variables, it is crucial to examine the correlation between the target and input variables. This allows for exploring the relationship between the model output, the school's electrical energy consumption, and the model input variables. Table 6 shows the input variables' correlations with the target energy consumption variable (kWh/year).

From the table, it is visible that:

- *TNU* has a correlation coefficient of 0,871359 with *AEC*, suggesting a strong positive correlation. This implies that changes in *TNU* are closely associated with changes in *AEC*, indicating *TNU*'s significant influence on electrical consumption.
- A_k exhibits a higher correlation coefficient of 0,944174 with *AEC*, indicating an even stronger positive correlation than *TNU*. This suggests that A_k has a particularly pronounced impact on *AEC*, potentially indicating its crucial role in determining electrical consumption.
- V_e also demonstrates a relatively high correlation coefficient of 0,896220 with *AEC*, indicating a strong positive correlation. This suggests that variations in V_e are closely linked to changes in *AEC*, highlighting V_e's importance in influencing electrical consumption.

Table 6
Correlations of input variables with target variable.

Variable	Correlation with AEC
TNU	0,870730
A_k	0,943930
V_e	0,895643

In this study, a crucial step was taken to ensure the reliability of our model. The initial data set was split into two sets using a random process: the training set, which was used to create or develop a model, and the validation set, which was used solely to validate the developed models' error assessment predictions. This study utilized original electrical energy usage data for 149 school buildings in Croatia's Osijek-Baranja County that was retrieved from EMIS. In the training set, 105 school buildings' worth of data were chosen at random, accounting for 70,5 % of the total of 149 school buildings. Additionally, there are 44 school buildings in the validation set or 29,5 % of the total. Similar ratios were also used in (Jain et al., 2014; Kontokosta and Tull, 2017).

Continuing with the paper, the results of developed models for predicting electrical energy consumption using multiple linear regression and neural networks are presented in the example of school buildings in Osijek-Baranja County in Croatia, using a given data set.

4. Developed models for predicting electrical energy consumption

4.1. MLR model

In order to identify which of all potential independent variables can affect the dependent variable, the regression models in this work are used to relate electrical energy usage to one or more variables. This procedure defines a mathematical model and establishes the relationships between the variables. Regression analysis makes it possible to define the relationship between independent and dependent variables, analyze variables regarding their availability, relevance, and collection, and construct mathematical models for prediction. Regression analysis is a technique for creating models that uses statistical analysis of essential variables and historical data (Ghania and Ahmad, 2010). In this study, the connection between one dependent and three independent variables was analyzed using MLR. The following is the typical form of the MLR equation with k independent variables (Papić, 2005):

$$\widehat{Y} = a + b_1 \cdot X_1 + b_2 \cdot X_2 + \ldots + b_k \cdot X_k \quad (1)$$

as follows:

- $\widehat{Y}$ expected or predicted value of the dependent variable,
- a intercept
- $b_{1\ldots k}$ regression coefficients and
- $X_{1\ldots n}$ of independent variable values.

In the development of the MLR model, a stepwise regression method was used, using the TIBCO Statistica® 14.1.0 (Cloud Software Group Inc, 2024) for the training set. A common data exploration tool called the stepwise method selects explanatory variables for an MLR model based on statistical significance (Smith, 2018). Each variable's p-value is determined throughout the stepwise regression process, and if a variable's p-value is higher than 0.05, it is eliminated from consideration until the best model with a consistent variable and a p-value of less than or equal to 0.05 is found (Alqahtani and Whyte, 2016).

The best developed MLR model for the dependent variable *AEC* which predicts the average annual electrical energy consumption has the following form:

$$AEC = 378,31 + 22,95TNU + 7,65 \quad A_k + 0,52 \quad V_e \quad [kWh/year] \qquad (2)$$

as follows:

- *TNU* independent variable representing the total number of users including employees and pupils (number),
- A_k independent variable representing the total useful surface area of the building (m^2),
- V_e independent variable representing the heated volume of the building (m^3),

4.2. ANN model

A multilayer perceptron (MLP) type of ANN was used to form the model. It is a type of ANN with multiple layers of interconnected nodes (Rana et al., 2018). These nodes process information and learn from data by applying weights and activation functions. MLP training aims to reduce the discrepancies between the model-calculated values and the intended target values. The weights are adjusted to decrease errors if the network provides an incorrect response or if the errors exceed a predetermined threshold (Widrow and Lehr, 1990). Errors are decreased, increasing the likelihood that subsequent network responses will be accurate. The network shows datasets containing pairs of the desired target and input patterns successively during the learning process. An MLP's learning algorithm consists of two steps: one for forward propagation and the other for backward propagation (Park and Lek, 2016). Fig. 4 shows the architecture of the optimal selected model.

The architecture of MLP refers to the number of layers and the number of nodes in each layer. In this case, 3–5-1 indicates the following:

- 3: There is an input layer with 3 nodes. These nodes represent the features or variables that are fed into the network.

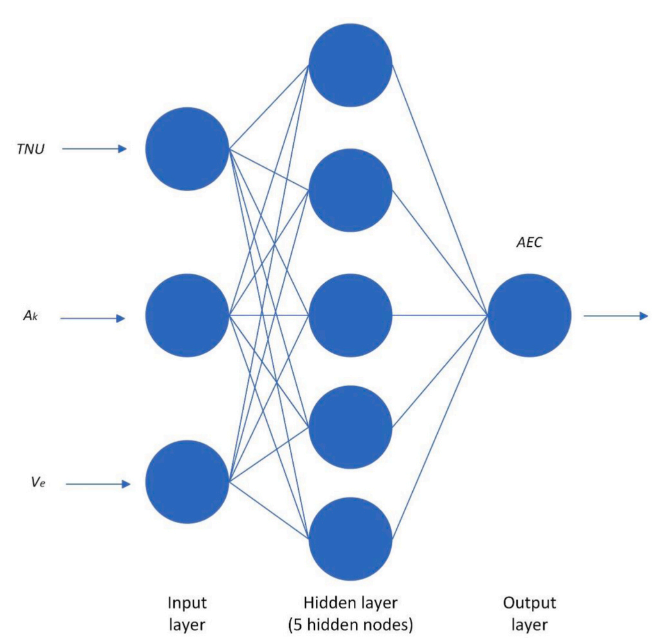

Fig. 4. Optimal selected model (MLP 3–5–1).

- 5: There is one hidden layer with 5 nodes. This layer performs the main information processing and learning within the network.
- 1: There is an output layer with 1 node. This node represents the network's prediction or output value.

It is also important to mention the stopping criteria where the early stopping and a maximum number of epochs were used to manage the training process. Specifically, training was stopped if there was no improvement in validation loss over 10 consecutive epochs, and the training at 100 epochs to avoid excessive computation. Besides, the control parameters included a learning rate of 0001, a batch size of 32, and the Adam optimizer. The network's hidden layer utilized the ReLU activation function, and L2 regularization with a penalty of 0,01 was applied to prevent overfitting. The MLP architecture consisted of an input layer with 3 nodes, a single hidden layer with 5 nodes, and an output layer with 1 node as previously mentioned.

5. Results

Finding the actual values of the expected (predicted) model outcomes—that is the degree to which the values of the dependent variables may be predicted—is essential (Hawkins, 2004). Representativeness, or the model's capacity to use a subset of independent variables to explain changes in the dependent variable, is assessed using absolute and relative metrics. These indicators are derived from the distribution of the dependent variable's value divergence from its anticipated and arithmetic means (McKelvey and Zavoina, 1975). To evaluate the accuracy of the produced prediction model and enable comparison with other models with various parameters, various statistical techniques are employed to assess the prediction error.

The following coefficients were used to estimate the prediction error of the developed models: mean absolute percentage error (MAPE), coefficient of determination (R^2), mean square error (MSE), root mean square error (RMSE), and coefficient of variation of the root mean square error (CVRMSE). Unlike most error metrics, MAPE (Mean Absolute Percentage Error) doesn't have a strict upper limit. The reason for this is because it involves taking the absolute value of the percentage error. Unlike most error metrics, MAPE doesn't have a strict upper limit. The reason for this is because it involves taking the absolute value of the percentage error (Saigal and Mehrotra, 2012). The closer the coefficient of determination R^2 value is to 1, the more representative the prediction model is (Schneider et al., 2010). Generally, a lower MSE is desirable as it signifies that the model's predictions are, on average, closer to the actual values. Conversely, a high MSE suggests that the model's predictions are consistently far off from the real values (Wang and Bovik, 2009). Typically, the upper limit for CVRMSE of 30 % is used to measure representativeness (Lulić, 2014). The equations utilized to compute the statistical techniques for prediction error estimation are displayed in Table 7.

Table 7
Expressions for the calculation of statistical methods for estimating prediction error.

No.	Coefficient	Expression	Ref.
1	R^2	$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{t=1}^{n}(y_i - \overline{Y})^2}$	(Papić, 2005)
2	MSE	$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_i - \hat{y}_i)^2$	(Sobol, 1991)
3	RMSE	$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_i - \hat{y})^2}$	(Sailee, 2019)
4	CVRMSE	$CVRMSE = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_i - \hat{y})^2}}{\overline{Y}} \cdot 100(\%)$	(Cacabelos et al., 2017)
5	MAPE	$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left\lvert\frac{(y_i - \hat{y}_i)}{y_i}\right\rvert \cdot 100(\%)$	(Small and Wong, 2002)

where it is as follows:

- y_i real values of the dependent variable,
- $\hat{y}_i$ predicted or expected values of the dependent variable and
- $\overline{Y}$ arithmetic mean of the dependent variable.

In this perspective, Table 8 presentes the values of the aforementioned coefficients for estimating the prediction error of the developed MLR model and the ANN model in the model training set.

It can be seen from the table that both MLR and ANN achieve high R^2 values (0950 and 0957 respectively) which indicates a strong positive correlation between the predicted and actual AEC values. However, the MLR model has a higher MSE compared to the ANN model. While MSE itself is difficult to interpret directly due to its units being squared errors, a lower MSE generally suggests a better fit. Similarly, the MLR model has a slightly higher RMSE compared to the ANN model. Following the same logic, a lower RMSE indicates the model's predictions are on average closer to the actual AEC values. Both models have very similar MAPE values (around 27 %).

With a CVRMSE of 24,99 the NN model appears to generalize slightly better to unseen data compared to the MLR model (CVRMSE of 26,83). This suggests that the NN model is less prone to overfitting the training data and might perform better on new data it hasn't been explicitly trained on.

Considering both $R^2$and CVRMSE, the ANN model seems to be the better choice for predicting AEC in this case. It achieves a strong correlation with the actual values while also generalizing slightly better to unseen data.

After the training set, the performances of the developed models were also verified on the validation set which is an independent set on which the model development was not performed. Table 9 presentes the values of the aforementioned coefficients for estimating the prediction error of the developed MLR model and the ANN model in the model validation set.

Both models achieve high R^2 (0.949 and 0.954) on the validation data, indicating a strong positive correlation between the predicted and actual AEC values. This suggests both models generalize reasonably well to unseen data. The ANN model maintains a slight edge in terms of MSE and RMSE compared to MLR. This suggests the ANN model's predictions are, on average, slightly closer to the actual AEC values in the validation set. Both models have very similar MAPE (Mean Absolute Percentage Error) values (around 24.5 %). Both MLR and ANN seem to perform well on the validation data, with the NN model having a slight advantage in terms of accuracy. The high R^2 values and similar MAPE suggest both models capture the overall trends in the data.

6. Discussion

The literature review showed that although much research has been conducted globally, there is a lack of research regarding predicting energy consumption in Croatia. Also, the review revealed a need to provide a straightforward and repeatable framework and model for estimating the energy consumption of school buildings. Therefore, this paper presented a study conducted in Osijek-Baranja County in Croatia to identify the variables influencing electrical energy consumption in school buildings and develop models for successful prediction of electrical energy consumption.

This study provides a comprehensive analysis of electrical energy consumption in Croatian school buildings by addressing two key research questions: identifying the key factors influencing electrical energy consumption and comparing the effectiveness of MLR and ANNs in predicting this consumption.

1. What are the key factors influencing electrical energy consumption in schools of Osijek-baranja county?

The research identified several crucial factors that influence electrical energy consumption in Croatian school buildings, which are significant for both understanding and managing energy use:

- Total Number of Users (Employees and Pupils): The number of individuals using the building, including both staff and students, has a substantial impact on energy consumption. Schools with higher occupancy levels require more energy for lighting, cooling, heating, and other electrical needs. This finding aligns with the general observation that increased building occupancy often correlates with higher energy use.
- Total Useful Surface Area: The overall floor area of the school building directly affects its energy consumption. Larger surface areas necessitate more lighting and potentially more extensive heating and cooling systems to maintain comfortable indoor conditions. This factor is consistent with global trends where larger building spaces typically lead to higher energy demands.
- Heated Volume of the Building: The volume of the building that needs to be heated is another critical factor. Schools with larger heated volumes generally consume more energy for heating purposes. This factor highlights the importance of building insulation and HVAC system efficiency in managing energy use, as buildings with greater volumes require more energy to achieve and maintain desired temperatures.

2. How do MLR and ANNs perform in predicting electrical energy consumption in these schools?

The comparative analysis of MLR and ANN models provides valuable insights into their predictive capabilities:

Performance Metrics: The ANN model consistently outperformed the MLR model in terms of accuracy and error metrics. With R^2 values of 0.957 for the training set and 0.954 for the validation set, the ANN model demonstrated a strong correlation with actual energy consumption data. Its RMSE and MAPE values also indicated superior predictive accuracy and generalizability compared to the MLR model. This performance is in line with previous research highlighting the efficacy of ANNs in energy prediction tasks.

Model Robustness: Both models showed a slight decrease in R^2 and an increase in error metrics when applied to the validation data, suggesting that neither model severely overfits the training data. However, the ANN model exhibited better performance in generalizing to new data, with lower CVRMSE values indicating a slightly reduced tendency for overfitting.

Comparative Effectiveness: The MLR model, while performing well, did not match the ANN model's accuracy. MLR achieved R^2 values of 0.950 and 0.949 for the training and validation sets, respectively, with slightly higher RMSE and MAPE values.

In conclusion, the findings suggest that while both MLR and ANN models are effective tools for predicting energy consumption, the ANN model offers a slight advantage in terms of accuracy and generalizability. This aligns with existing literature and reinforces the value of

Table 8
Statistical analysis of the MLR and ANN models' prediction errors in the training data set.

No.	Dependent variable	Model type	R^2	MSE	RMSE	CVRMSE	MAPE
1	AEC	MLR	0950	1,054E+07	3246,19	26,83 %	27,70 %
2	AEC	NN	0957	9,146E+05	3024,25	24,99 %	27,28 %

Table 9
Statistical analysis of the MLR and ANN models' prediction errors in the validation data set.

No.	Dependent variable	Model type	R^2	MSE	RMSE	CVRMSE	MAPE
1	AEC	MLR	0949	1,251E+07	3537,51	20,50 %	24,60 %
2	AEC	NN	0954	1,167E+07	3415,75	19,79 %	24,53 %

employing advanced predictive modeling techniques for energy management in school buildings. Future research should continue to explore these methodologies and consider integrating additional factors and data sources to further enhance prediction accuracy and efficiency. As mentioned above, the ANN model, configured with a 3–5-1 architecture, achieved R^2 values of 0957 and 0954 for the training and validation datasets, respectively, indicating a strong correlation between predicted and actual values. The RMSE values of 3024,25 in the training set and 3415,75 in the validation set, coupled with MAPE values of 27,28 % and 24,53 %, suggest that the ANN model provides accurate predictions and generalizes effectively to unseen data. This performance is comparable to the findings of González and Zamarreño (Gonzalez and Zamarreno, 2005), who reported high accuracy in electric load forecasting using a feedback ANN, and Ahmad et al (Ahmad et al., 2017)., who observed a slightly better performance of ANN over random forests in predicting hotel HVAC energy consumption.

In comparison, as mentioned earlier, the MLR model achieved R^2 values of 0950 and 0949 for the training and validation sets, respectively, with RMSE values of 3246,19 and 3537,51. The MAPE values of 27,70 % and 24,60 % are consistent with the performance of similar models in the literature. For instance, Tso and Yau's (Tso and Yau, 2007) regression analysis model, while comparable to decision tree and ANN models, demonstrated slight variations in performance across seasons. Similarly, the MLR model proposed by Run et al. for hourly electricity usage in school buildings achieved lower R^2 values of 74 % and 77 % for training and testing, respectively (Run et al., 2023). However, the current study's MLR model demonstrates superior predictive accuracy. Overall, the models developed in this study exhibit strong performance metrics, with the ANN model showing a slight edge in generalization, making it a reliable tool for predicting energy consumption in buildings.

7. Conclusion

The ANN model exhibited superior accuracy and generalizability in predicting energy consumption, making it a powerful tool for this purpose. However, this increased performance comes at the cost of interpretability, posing significant challenges for non-technical users. The intricate architecture of neural networks, characterized by numerous interconnected layers and weights, complicates the understanding of how input variables influence output predictions. This complexity can be a barrier for users who need to trust and comprehend the model's decisions, particularly in contexts where transparency is crucial.

Conversely, the MLR model, while slightly less accurate than the ANN model, offers the advantage of simplicity and ease of interpretation. Its straightforward nature makes it more accessible to users without advanced technical expertise, allowing them to easily understand how different building characteristics affect energy use. Given that the model is intended for use by school facility managers or principals—who may lack specialized knowledge in machine learning—the MLR model's transparency becomes a critical asset. Although the ANN model outperforms the MLR in predictive performance, its complexity could hinder its adoption in practical settings.Therefore, despite its lower accuracy, the MLR model may be more suitable for facilitating informed decision-making in school energy management.

However, this study has several limitations. First, the models were developed and validated using data from a specific county, which may limit their generalizability to other regions with different climatic conditions and building practices. Yet, while the current model is optimized for a specific county, its underlying framework provides a robust foundation for adaptation. Therefore, as a future research direction, the authors propose that future studies should validate the model in various geographical settings by collecting data from schools in different regions and considering their unique characteristics. However, since Croatia is divided into continental and coastal regions, this model ensures applicability to the whole continental part of Croatia since cities and towns with 2200 and more degree days of heating and annual energy needs are calculated according to the reference climate data for continental Croatia. Also, it is essential to note that the study accounted for diverse school-building practices by including both primary and secondary schools, enhancing the model's robustness.

The complexity of the ANN model also presents a limitation, as it may require significant computational resources and expertise to implement and interpret effectively. Therefore, as a future research direction, it would be beneficial to explore the integration of advanced computational techniques such as quantum computing, which holds promise in optimizing model performance and handling complex datasets more efficiently. Additionally, adapting convolutional neural networks (CNNs) enables the investigation of spatial and temporal data in energy consumption dynamics within school facilities. CNNs have shown effectiveness in pattern recognition tasks and could offer new perspectives on identifying energy usage patterns that traditional models may overlook.

As another limitation, the study did not explore the long-term impact of energy savings by using these predictive models, leaving room for further investigation. In this perspective, as a future research direction it is proposed to extend the study by applying the developed predictive models in a long term period, such as several years, to analyse the effects of using the models on actual consumption reduction.

To conclude, the significance of this study lies in its contribution to the field of energy management in educational institutions. By comparing the accuracy and interpretability of the ANN and MLR models, this research provides valuable insights into the trade-offs between model complexity and usability. The findings underscore the importance of considering the end-user's technical expertise when selecting predictive models for practical applications. Moreover, the proposed future research directions offer a pathway for enhancing the models' robustness, generalizability, and practical utility, ultimately contributing to more effective energy management strategies in schools.

CRediT authorship contribution statement

Hrvoje Krstić: Writing – review & editing, Validation, Supervision, Methodology, Formal analysis. **Hana Begić:** Writing – original draft, Visualization, Investigation, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Afzal, S., Shokri, A., Ziapour, B.M., Shakibi, H., Sobhani, B., 2024. Building energy consumption prediction and optimization using different neural network-assisted models; comparison of different networks and optimization algorithms. Eng. Appl. Artif. Intell. 127, 107356.

Agencija za pravni promet i posredovanje nekretninama, 2015. Inf. Sustav-.-. za Gospod. Energ. - Isge. Available from: ⟨https://apn.hr/gospodarenje-energijom-isge/informacijski-sustav-za-gospodarenje-energijom⟩.

Ahmad, M.W., Mourshed, M., Rezgui, Y., 2017. Trees vs Neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy Build. 147, 77–89. https://doi.org/10.1016/j.enbuild.2017.04.038.

Alghoul, S.K., Gwesha, A.O., Naas, A.M., 2016. The effect of electricity price on saving energy transmitted from external building walls. Energy Res. J. 7 (1), 1–9. https://doi.org/10.3844/erjsp.2016.1.9.

Alqahtani, A., Whyte, A., 2016. Estimation of life-cycle costs of buildings: regression vs artificial neural network. Built Environ. Proj. Asset Manag. 6 (1), 30–43. https://doi.org/10.1108/BEPAM-08-2014-0035.

Alshibani, A., 2020. Prediction of the energy consumption of school buildings. Appl. Sci. 10 (17), 5885. https://doi.org/10.3390/app10175885.

Araújo, I., Nunes, L.J., Curado, A., 2023. Preliminary approach for the development of sustainable university campuses: a case study based on the mitigation of greenhouse gas emissions. Sustainability 15 (6), 5518.

Bećirović, S.P., Vasić, M., 2013. Methodology and results of Serbian energy-efficiency refurbishment project. Energy Build. 62, 258–267. https://doi.org/10.1016/j.enbuild.2013.03.027.

Bertone, E., Sahin, O., Stewart, R.A., Zou, P.X., Alam, M., Hampson, K., Blair, E., 2018. Role of financial mechanisms for accelerating the rate of water and energy efficiency retrofits in Australian public buildings: hybrid Bayesian network and system dynamics modelling approach. Appl. Energy 210, 409–419.

Beusker, E., Stoy, C., Pollalis, S.N., 2012. Estimation model and benchmarks for heating energy consumption of schools and sport facilities in Germany. Build. Environ. 49, 324–335. https://doi.org/10.1016/j.buildenv.2011.08.006.

Bhowmik, C., Bhowmik, S., Ray, A., Pandey, K.M., 2017. Optimal green energy planning for sustainable development: a review. Renew. Sustain. Energy Rev. 71, 796–813. https://doi.org/10.1016/j.rser.2016.12.10.

Biswas, M.R., Robinson, M.D., Fumo, N., 2016. Prediction of residential building energy consumption: a neural network approach. Energy 117, 84–92. https://doi.org/10.1016/j.energy.2016.10.066.

Børke, R., 2006. Energy efficiency in non-residential buildings: Motivation, barriers and strategies, in Program for industriell økologi. NTNU: Norwegian University of Science and Technology, Trondheim, Norway.

Cacabelos, A., Eguía, P., Febrero, L., Granada, E., 2017. Development of a new multi-stage building energy model calibrationmethodology and validation in a public library. Energy Build. 146, 182–199. https://doi.org/10.1016/j.enbuild.2017.04.071.

Cao, X., Dai, X., Liu, J., 2016. Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade. Energy Build. 128, 198–213. https://doi.org/10.1016/j.enbuild.2016.06.089.

Cao, W., Yu, J., Chao, M., Wang, J., Yang, S., Zhou, M., Wang, M., 2023. Short-term energy consumption prediction method for educational buildings based on model integration. Energy 283, 128580.

Capozzoli, A., Grassi, D., Causone, F., 2015. Estimation models of heating energy consumption in schools for local authorities planning. Energy Build. 105, 302–313. https://doi.org/10.1016/j.enbuild.2015.07.024.

Chen, L., Chen, Z., Zhang, Y., Liu, Y., Osman, A.I., Farghali, M., Hua, J., Al-Fatesh, A., Ihara, I., Rooney, D.W., 2023. Artificial intelligence-based solutions for climate change: a review. Environ. Chem. Lett. 21 (5), 2525–2557. https://doi.org/10.1007/s10311-023-01617-y.

Cloud Software Group Inc. TIBCO Statistica® 14.1.0. 2024; Available from: ⟨https://docs.tibco.com/products/tibco-statistica-14-1-0⟩.

Corgnati, S.P., Corrado, V., Filippi, M., 2008. A method for heating consumption assessment in existing buildings: a field survey concerning 120 Italian schools. Energy Build. 40 (5), 801–809. https://doi.org/10.1016/j.enbuild.2007.05.011.

D'Agostino, D., Cuniberti, B., Bertoldi, P., 2017. Energy consumption and efficiency technology measures in European non-residential buildings. Energy Build. 153, 72–86. https://doi.org/10.1016/j.enbuild.2017.07.062.

Deng, Z., Wang, X., Dong, B., 2023. Quantum computing for future real-time building HVAC controls. Appl. Energy 334, 120621.

Desideri, U., Proietti, S., 2002. Analysis of energy consumption in the high schools of a province in central Italy. Energy Build. 34 (10), 1003–1016. https://doi.org/10.1016/S0378-7788(02)00025-7.

Dimoudi, A., Kostarela, P., 2009. Energy monitoring and conservation potential in school buildings in the C′ climatic zone of Greece. Renew. Energy 34 (1), 289–296.

Doiphode, G., Najafi, H., 2020. A machine learning based approach for energy consumption forecasting in K-12 Schools. ASME International Mechanical Engineering Congress and Exposition. American Society of Mechanical Engineers.

Đukanović, L., Ignjatović, D., Ćuković Ignjatović, N., Rajčić, A., Lukić, N., Zeković, B., 2022. Energy refurbishment of Serbian school building stock—A typology tool methodology development. Sustainability 14 (7), 4074. https://doi.org/10.3390/su14074074.

El Alaoui, M., Chahidi, L.O., Rougui, M., Lamrani, A., Mechaqrane, A., 2023. Prediction of energy consumption of an administrative building using machine learning and statistical methods. Civ. Eng. J. 9 (5), 1007–1022.

Elhabyb, K., Baina, A., Bellafkih, M., Deifalla, A.F., 2024. Machine learning algorithms for predicting energy consumption in educational buildings. Int. J. Energy Res. 2024 (1), 6812425.

Faiq, M., Tan, K.G., Liew, C.P., Hossain, F., Tso, C.-P., Lim, L.L., Wong, A.Y.K., Shah, Z. M., 2023. Prediction of energy consumption in campus buildings using long short-term memory. Alex. Eng. J. 67, 65–76.

Gaitani, N., Lehmann, C., Santamouris, M., Mihalakakou, G., Patargias, P., 2010. Using principal component and cluster analysis in the heating evaluation of the school building sector. Appl. Energy 87 (6), 2079–2086. https://doi.org/10.1016/j.apenergy.2009.12.007.

Geraldi, M.S., Ghisi, E., 2020. Building-level and stock-level in contrast: a literature review of the energy performance of buildings during the operational stage. Energy Build. 211, 109810. https://doi.org/10.1016/j.enbuild.2020.109810.

Ghania, I.M.M., Ahmad, S., 2010. Stepwise Multiple Regression Method to Forecast Fish Landing. In: Tarmizi, R.A. (Ed.), International Conference on Mathematics Education Research 2010 (ICMER 2010). Elsevier Procedia, Malacca, Malaysia, pp. 549–554. https://doi.org/10.1016/j.sbspro.2010.12.076.

Giani, A., Eldredge, Z., 2021. Quantum computing opportunities in renewable energy. SN Comput. Sci. 2 (5), 393.

Gonzalez, P.A., Zamarreno, J.M., 2005. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. Energy Build. 37 (6), 595–601. https://doi.org/10.1016/j.enbuild.2004.09.006.

Hawkins, D.M., 2004. The problem of overfitting. J. Chem. Inf. Comput. Sci. 44 (1), 1–12. https://doi.org/10.1021/ci0342472.

Issa, M., Attalla, M., Rankin, J., Christian, A., 2010. Detailed analysis of electricity, water, and gas consumption quantities and costs in Toronto's public schools. Can. J. Civ. Eng. 37 (1), 25–36. https://doi.org/10.1139/L09-122.

Jain, R.K., Smith, K.M., Culligan, P.J., Taylor, J.E., 2014. Forecasting energy consumption of multi-family residential buildings using support vector regression: investigating the impact of temporal and spatial monitoring granularity on performance accuracy. Appl. Energy 123, 168–178. https://doi.org/10.1016/j.apenergy.2014.02.05.

Jovanović, M.P., Vučićević, B.S., Turanjanin, V.M., Lazović, I.M., Živković, M.M., 2018. Assessing the sustainability of Serbian school buildings by analyse and syntesis parameters under information deficiency method. Therm. Sci. 22, 1271–1283. https://doi.org/10.2298/TSCI170529131J.

Jurišević, N., Gordić, D., Lukić, N., Josijević, M., 2018. Benchmarking heat consumption in educational buildings in the city of Kragujevac (Serbia). Energy Effic. 11, 1023–1039.

Katafygiotou, M., Serghides, D.K., 2014. Analysis of structural elements and energy consumption of school building stock in Cyprus: Energy simulations and upgrade scenarios of a typical school. Energy Build. 72, 8–16. https://doi.org/10.1016/j.enbuild.2013.12.024.

Katic, D., Krstic, H., Marenjak, S., 2021. Energy Performance of School Buildings by Construction Periods in Federation of Bosnia and Herzegovina. Buildings 2021 11, 42.s Note: MDPI stays neutral with regard to jurisdictional claims in published ….

Katić, D., Krstić, H., Marenjak, S., 2020. U-values of school building envelopes in the south region of the Federation of Bosnia and Herzegovina. e-ZBORNIK, Electron. Collect. Pap. Fac. Civ. Eng. 10 (20), 55–67. https://doi.org/10.47960/2232-9080.2020.20.10.57.

Katić, D., Krstić, H., 2022. Energy consumption for heating of school buildings in the south region of the Federation of Bosnia and Herzegovina. e-ZBORNIK Electron. Collect. Pap. Fac. Civ. Eng. 12 (23). https://doi.org/10.47960/2232-9080.2022.23.12.20.

Katić, D., Krstić, H., Marenjak, S., 2021. Energy performance of school buildings by construction periods in federation of Bosnia and Herzegovina. Buildings 11 (2), 42.

Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z., Djurovic-Petrovic, M., 2010. A review of bottom-up building stock models for energy consumption in the residential sector. Build. Environ. 45 (7), 1683–1697. https://doi.org/10.1016/j.buildenv.2010.01.021.

Kim, T., Kang, B., Kim, H., Park, C., Hong, W.-H., 2019. The study on the energy consumption of middle school facilities in Daegu, Korea. Energy Rep. 5, 993–1000. https://doi.org/10.1016/j.egyr.2019.07.015.

Kontokosta, C.E., Tull, C., 2017. A data-driven predictive model of city-scale energy use in buildings. Appl. Energy 197, 303–317. https://doi.org/10.1016/j.apenergy.2017.04.005.

Krstić, H., Teni, M., 2018. Analysis of energy performance and buildings characteristics obtained from Croatian energy management information system. Int. J. Struct. Civ. Eng. Res 7 (3), 252–258. https://doi.org/10.18178/ijscer.7.3.252-258.

te Kulve, M., Hellwig, R.T., van Dijken, F., Boerstra, A., 2022. Do children feel warmer than adults? Overheating prevention in schools in the face of climate change. Routledge Handbook of Resilient Thermal Comfort. Routledge, pp. 128–140.

Li, X., Chen, S., Li, H., Lou, Y., Li, J., 2023. A behavior-orientated prediction method for short-term energy consumption of air-conditioning systems in buildings blocks. Energy 263, 125940.

Li, C., Ding, Z., Zhao, D., Yi, J., Zhang, G., 2017. Building energy consumption prediction: an extreme deep learning approach. Energies 10 (10), 1525. https://doi.org/10.3390/en10101525.

Lizana, J., Serrano-Jimenez, A., Ortiz, C., Becerra, J.A., Chacartegui, R., 2018. Energy assessment method towards low-carbon energy schools. Energy 159, 310–326.

Lulić, I., 2014. The use of regression analysis method in solving problems from engineering practice. Faculty of Mechanical Engineering and Naval Architecture. University of Zagreb, Zagreb, Croatia.

McKelvey, R.D., Zavoina, W., 1975. A statistical model for the analysis of ordinal level dependent variables. J. Math. Sociol. 4 (1), 103–120. https://doi.org/10.1080/0022250X.1975.9989847.

Microsoft. Microsoft Excel. 2024; Available from: ⟨https://www.microsoft.com/en-us/microsoft-365/excel⟩.
Ministarstvo graditeljstva i prostornoga uređenja, Pravilnik o sustavnom gospodarenju energijom u javnom sektoru (NN 18/2015) 2015.
Ministarstvo prostornog uređenja, g.i.d.i.R.H., Long-Term Strategy for National Building Stock Renovation by 2050 2020.
Mohammed, A., Alshibani, A., Alshamrani, O., Hassanain, M., 2021. A regression-based model for estimating the energy consumption of school facilities in Saudi Arabia. Energy Build. 237, 110809. https://doi.org/10.1016/j.enbuild.2021.110809.
Nutakki, M., Koduru, S., Mandava, S., 2024. Quantum support vector machine for forecasting house energy consumption: a comparative study with deep learning models. J. Cloud Comput. 13 (1), 1–12.
Olu-Ajayi, R., Alaka, H., Owolabi, H., Akanbi, L., Ganiyu, S., 2023. Data-driven tools for building energy consumption prediction: a review. Energies 16 (6), 2574. https://doi.org/10.3390/en16062574.
Papadakis, N., Katsaprakakis, D.A., 2023. A review of energy efficiency interventions in public buildings. Energies 16 (17), 6329.
Papić, M., 2005. Primijenjena statistika u MS Excelu za ekonomiste, znanstvenike i neznalice. Zoro d.o.o, Zagreb, Croatia.
Park, Y.-S., Lek, S., 2016. Artificial neural networks: Multilayer perceptron for ecological modeling. Developments in environmental modelling. Elsevier, pp. 123–140. https://doi.org/10.1016/B978-0-444-63623-2.00007-4.
Pereira, L.D., Raimondo, D., Corgnati, S.P., Da Silva, M.G., 2014. Energy consumption in schools–A review paper. Renew. Sustain. Energy Rev. 40, 911–922. https://doi.org/10.1016/j.rser.2014.08.010.
Perez, Y.V., Capeluto, I.G., 2009. Climatic considerations in school building design in the hot–humid climate for reducing energy consumption. Appl. Energy 86 (3), 340–348. https://doi.org/10.1016/j.apenergy.2008.05.007.
Raatikainen, M., Skön, J.-P., Leiviskä, K., Kolehmainen, M., 2016. Intelligent analysis of energy consumption in school buildings. Appl. Energy 165, 416–429. https://doi.org/10.1016/j.apenergy.2015.12.072.
Rana, A., Rawat, A.S., Bijalwan, A., Bahuguna, H., 2018. Application of multi layer (perceptron) artificial neural network in the diagnosis system: A systematic review. 2018 International conference on research in intelligent and computing in engineering (RICE). IEEE, San Salvador, El Salvador. https://doi.org/10.1109/RICE.2018.8509069.
Raza, M.Q., Khosravi, A., 2015. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renew. Sustain. Energy Rev. 50, 1352–1372. https://doi.org/10.1016/j.rser.2015.04.065.
Republic of croatia ministry of economy and sustainable development, 2022. Energy Croat. 2022.
Run, K., Cévaër, F., Dubé, J.-F., 2023. Preliminary multiple linear regression model to predict hourly electricity consumption of school buildings. Future Energy: Challenge, Opportunity, and, Sustainability. Springer, pp. 119–127.
Saigal, S., Mehrotra, D., 2012. Performance comparison of time series data using predictive data mining techniques. Adv. Inf. Min. 4 (1), 57–66.
Sailee, R., 2019. Exploration of Variable Importance and Variable selection techniques in presence of correlated variables. Department of Mathematical Sciences. Rochester Institute of Technology, College of Science, Rochester, New York.
Schneider, A., Hommel, G., Blettner, M., 2010. Linear regression analysis: part 14 of a series on evaluation of scientific publications. Dtsch. Ärzteblatt Int. 107 (44), 776. https://doi.org/10.3238/arztebl.2010.0776.
Schober, P., Boer, C., Schwarte, L.A., 2018. Correlation coefficients: appropriate use and interpretation. Anesth. Analg. 126 (5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864.
Shahid, Z.K., Saguna, S., Åhlund, C., 2023. Forecasting electricity and district heating consumption: A case study in schools in Sweden. 2023 IEEE Green Technologies Conference (GreenTech). IEEE.
Small, G.R., Wong, R., 2002. The Validity of Forecasting. In: Parker, D. (Ed.), The 8th Annual PRRES Conference in Christchurch, New Zealand. PRRES Inc, Christchurch, New Zealand, pp. 1–14.
Smith, G., 2018. Step away from stepwise. J. Big Data 5 (1), 1–12. https://doi.org/10.1186/s40537-018-0143-6.
Sobol, M.G., 1991. Validation strategies for multiple regression analysis: using the coefficient of determination. Interfaces 21, 106–120. https://doi.org/10.1287/inte.21.6.106.
Stanković, S., Campbell, N., Maksimović, D., Cvjetković, T., 2009. Evaluation of energy efficiency measures applied in public buildings (schools & hospitals) in Serbia. Spatium (20), 1–8. https://doi.org/10.2298/SPAT0920001S.
Tariq, R., Mohammed, A., Alshibani, A., Ramírez-Montoya, M.S., 2024. Complex artificial intelligence models for energy sustainability in educational buildings. Sci. Rep. 14 (1), 15020.
Teli, D., Bourikas, L., James, P.A., Bahaj, A.S., 2017. Thermal performance evaluation of school buildings using a children-based adaptive comfort model. Procedia Environ. Sci. 38, 844–851.
Tso, G.K., Yau, K.K., 2007. Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks. Energy 32 (9), 1761–1768. https://doi.org/10.1016/j.energy.2006.11.010.
Vassallo, P., 2020. Analysing the" performance gap" between energy performance certificates and actual energy consumption of non-residential buildings in Malta, in Institute for Sustainable Energy. University of Malta, Marsaxlokk, Malta.
Wang, Z., Bovik, A.C., 2009. Mean squared error: love it or leave it? A new look at signal fidelity measures. IEEE Signal Process. Mag. 26 (1), 98–117. https://doi.org/10.1109/MSP.2008.930649.
Wang, L., Xie, D., Zhou, L., Zhang, Z., 2023. Application of the hybrid neural network model for energy consumption prediction of office buildings. J. Build. Eng. 72, 106503.
Widrow, B., Lehr, M.A., 1990. 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. Proc. IEEE 78 (9), 1415–1442. https://doi.org/10.1109/5.58323.
Xu, F., Liu, Q., 2023. Building energy consumption optimization method based on convolutional neural network and BIM. Alex. Eng. J. 77, 407–417.
Zapata-Lancaster, M.G., Ionas, M., Toyinbo, O., Smith, T.A., 2023. Carbon dioxide concentration levels and thermal comfort in primary school classrooms: What pupils and teachers do. Sustainability 15 (6), 4803.